Original Article • Journal of Southeast Asian Orthopaedics Vol 50 No 2 (2026) 5-11

Journal of Southeast Asian Orthopaedics
ISSN 2821-9848 (Print)
ISSN 2821-9864 (Online)
https://doi.org/10.56929/jseaortho-2026-0287 https://jseaortho.org

The Effectiveness of a Machine Learning Model in Predicting Blood Transfusion Probability in Bipolar Hemiarthroplasty Hip Replacement Surgery

Polasan Santanapipatkul, MD [1], Anchalee Chumjam, MD [2], Lakkana Jirapong, MD [3]

[1] *Department of Orthopedic Surgery, Samut Sakhon Hospital, Samut Sakhon, Thailand*
[2] *Department of Otolaryngology, Samut Sakhon Hospital, Samut Sakhon, Thailand*
[3] *Department of Radiology, Samut Sakhon Hospital, Samut Sakhon, Thailand*

Purpose: To verify a machine learning-based prediction model for blood transfusion risk in patients undergoing bipolar hemiarthroplasty and to determine whether there are significant differences between the accuracy results of this verification and the original model.
Methods: A retrospective study using purposive sampling was designed to gather 136 samples with the inclusion criterion of undergoing bipolar hemiarthroplasty for femoral neck fractures at the author's institution between January 1, 2021, and June 30, 2024. The research instruments included (1) a machine learning-based prediction model for blood transfusion probability (smskbl.streamlit.app), which was constructed using 232 femoral neck fracture samples undergoing bipolar hemiarthroplasty at the author's institution from 2015 to 2020, and (2) a research questionnaire created by the researcher, including six items: one on demographic data, four on medical health conditions, and one on actual blood transfusion during surgery.
Results: The prediction model accuracy was 89%, compared with that of the original model (80%). The comparison of the accuracy results was not statistically significant ($Z = 0.424$, $p > 0.05$). In the blood transfusion group, the precision was 0.70, recall was 0.73, and F1-score was 0.72, whereas the group that did not receive blood transfusion had a precision of 0.94, recall of 0.93, and an F1-score of 0.93. The area under the curve was 0.83.
Conclusions: The blood transfusion prediction model demonstrated good performance in predicting transfusion risk. The model provides confidence in its risk prediction outcome and can be used to perform optimal risk management in preparation for bipolar hemiarthroplasty.

Keywords: machine learning, fracture neck of femur, bipolar hemiarthroplasty, transfusion

Article history:
Received: September 29, 2025 Revised: December 15, 2025
Accepted: January 29, 2026
Correspondence to: Polasan Santanapipatkul, MD
Department of Orthopedic Surgery, Samut Sakhon Hospital, Samut Sakhon, Thailand
E-mail: polasants@gmail.com

Hip replacement surgery (bipolar hemiarthroplasty), one of the most common orthopedic surgeries, is typically performed for femoral neck fractures[(1)]. Even simple surgical procedures can lead to severe hemorrhage depending on several factors. These factors include underlying medical conditions (e.g., bleeding disorders or severity exacerbated by anticoagulant medication), prolonged operative time, and increased intraoperative blood loss(2-3). To prevent life-threatening situations during surgery, a routine blood request is normally prepared to ensure that transfusion is needed(4). Typically, the amount of blood requested is based on agreement between the surgical team and the anesthesiologist, and is mostly influenced by the patient's risk factors. These factors include a medical history of heart or kidney disease, low preoperative hemoglobin or anemia(5-6), or a rare blood type in both the ABO type and Rh group(7). According to the American Society of Anesthesiologists (ASA) physical status classification, which ranges from Class I (a normal healthy patient) to Class VI (a declared brain-dead organ donor), a patient with a higher ASA score is more likely to require a larger amount of blood for surgery(8).

Currently, artificial intelligence (AI) is being acknowledged and rapidly implemented across many areas, especially in advanced diagnostics in healthcare(9). Experts agree that blood transfusion prediction models are crucial for determining whether a transfusion is required in each surgical case. All blood transfusion preparation procedures, such as blood crossmatch (2 units), one of the most important requirements, contribute to the cost of the operation in each case(10). Often, this blood preparation is not used because the patient remains in good condition during real-time monitoring. Patients maintained stable vital signs, and blood loss was minimal during surgery. In the author's institution during the years 2021–2024, the prevalence of blood transfusion in cases undergoing hemiarthroplasty was 19% (26 of 136 cases), indicating that the remaining 81% of the blood preparations were not used.

Several studies have developed blood transfusion prediction models based on machine learning (ML) for patients undergoing hip arthroplasty, including partial hip arthroplasty, which includes bipolar hemiarthroplasty. Many studies have shown that the final predictive models retained a wide range of risk factors. Liang et al.(11) identified 19 variables included in their predictive model. Buddhiraju et al.(12) developed a simple predictive blood transfusion model and included three main risk factors: (1) preoperative hemoglobin concentration, (2) hematocrit level, and (3) operative time. Most predictive models have been broadly used to improve the quality of care by decreasing the cost of service operations and eliminating blood supply waste, which is of high value to other patients. Several ML algorithms were used to develop these models. Some model outcomes showed superior performance; for example, Liang et al.(11) reported five models that showed superior performance with an area under the curve (AUC) value exceeding 0.90, including (1) logistic regression, (2) random forest (RF), (3) support vector machines (SVM), (4) K-nearest neighbors, and (5) naive Bayes (NB). RF was reported to yield the best results, with an accuracy of 0.86, precision of 0.80, specificity of 0.91, F1-score of 0.78, and sensitivity of 0.76.

In Thailand, some predictive models have been developed that include several risk factors such as low preoperative hemoglobin level, low body mass index, and the use of general anesthesia during surgery(13). However, despite the aim of improving healthcare quality, integrating these predictive models was difficult because of limited accessibility. Therefore, in this study, the previous developed model(14) between 2015 and 2020 from 232 femoral neck fracture samples that underwent bipolar hemiarthroplasty at the author's institution, was available. The model is constructed using AI based on ML concept, a crucial and well-known technique that allows computers to learn from historical data to forecast future trends without being explicitly programmed for every task(15). The model has been approved for clinical application and is currently available at smskbl.streamlit.app. The model comprised five main risk factors: one demographic factor (gender) and four other underlying medical conditions, including (1) chronic kidney disease, (2) ischemic heart disease, (3) prosthesis type, and (4) ASA classification score(13-14).

This model was used to predict the probability of blood transfusion in patients undergoing bipolar hemiarthroplasty. The predicted results were compared with the actual intraoperative transfusion data for each sample. The actual blood transfusion criterion was applied when the hematocrit threshold was <30% or <25% in patients with chronic anemia(16).

Purpose

1. To verify the ML-based risk prediction model for blood transfusion in patients undergoing bipolar hemiarthroplasty.

2. To determine whether there are significant differences between the accuracy results from this study and the original version reported by the author who created the tool.

METHODS

This retrospective study was designed to gather data from 136 femoral neck fracture samples that underwent bipolar hemiarthroplasty at the author's institution between January 1, 2021, and June 30, 2024. The inclusion criteria were patients with femoral neck fractures who underwent bipolar hemiarthroplasty. Patients for whom required personal risk factor data could not be collected were excluded.

Research Instruments

1. The ML-based predictive model of blood transfusion probability (smskbl.streamlit.app) was developed at the author's institution between 2015 and 2020 (5 years) using 232 femoral neck fracture samples that underwent bipolar hemiarthroplasty. To develop the model, 80% of the total samples were randomized using a computer system and were employed as the training set. Ten percent of the remaining samples were employed as the validation set, and the final 10% served as the test set to complete the model development. Three algorithm techniques—(1) NB, (2) SVM, and (3) RF—were employed. RF yielded the best conclusion, with an accuracy of 0.80.

2. A questionnaire created by the researcher included two demographic items (age and gender); three medical health condition items of binary yes/no choices (chronic kidney disease, ischemic heart disease, and cemented prosthesis); actual blood transfusion during surgery; and one item for the actual ASA physical classification score (I to VI) preoperatively.

Data Collection and Data Analysis

After obtaining approval from the ethics committee (SKH REC 111/2567/V.1), questionnaire data were retrieved from the author's institutional database (HOSxP). Blood transfusion was subsequently calculated using the authors' model to determine the probability of blood transfusion due to surgery. Data were analyzed using descriptive statistics to demonstrate the effectiveness of the tool. A *Z*-test was performed to compare the significant differences between two accuracy proportions: (1) the accuracy obtained in this study and (2) the previously reported accuracy of 0.80 from the original model.

RESULTS

Of the 136 samples, 106 (77.9%) were male and 30 (22.1%) were female (Table 1). Twenty-six patients (19.1%) actually received a blood transfusion during surgery, while the remaining 110 (80.9%) did not. After using the ML-based risk prediction model, only 19 of these 26 cases were correctly predicted as true positives (0.73), indicating that 19 samples were correctly identified as having received blood transfusion. In addition, 102 out of 110 cases were correctly predicted as true negatives (0.93), indicating that these samples did not require blood transfusion (Table 2). The accuracy of the prediction model was 89% (95% confidence interval: 83–94%) (Table 3). Comparison of the accuracy between the results of this study and the original model (0.89 vs. 0.80) was performed using a *Z*-test, which showed no statistically significant difference ($Z = 0.424$, $p > 0.05$).

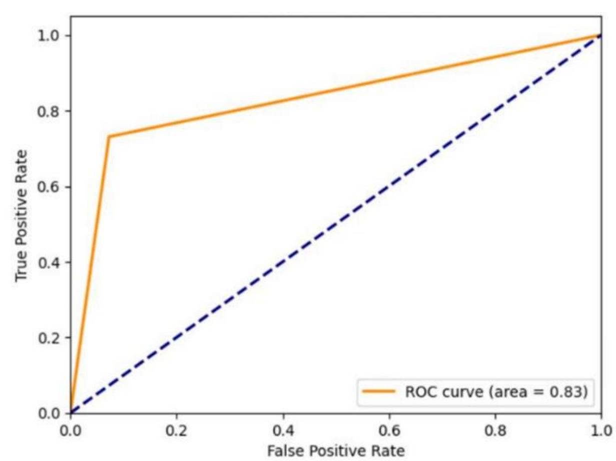

Fig. 1 The area under the curve of the blood transfusion prediction model.

In the group who did not receive blood transfusion, the screening tool statistics were as follows: precision = 0.94, recall = 0.93, and F1-score = 0.93. In the group who received blood transfusion, the screening tool statistics were precision = 0.70, recall = 0.73, and F1-score = 0.72 (Table 4). The AUC was 0.83 (Figure 1).

Table 1 Demographic data of samples. (N = 136)

Demographic data (Category)	Number	Percentage (%)
Gender (Female/Male)	30/106	22.1/77.9
Age Mean = 74.8 years (S.D. = 9.37)		
Minimum = 54, Maximum = 93		
Chronic Kidney Disease (No/Yes)	112/24	82.4/17.6
Ischemic Heart Disease HD (No/Yes)	119/17	87.5/12.5
Cemented Prosthesis (No/Yes)	124/12	91.2/8.8
ASA Classification (Class 2/3/4)	13/117/6	9.6/86.0/4.4
Actual Blood Transfusion (No/Yes)	110/26	80.9/19.1
Predicted Blood Transfusion risk (No/Yes)	109/27	80.1/19.9

Abbreviations: S.D., standard deviation; ASA, American Society of Anesthesiologists

Table 2 Logical model prediction outcome and actual blood transfusion condition.

Model Prediction	Actual blood transfusion		Total
	Yes	No	
Yes	19 (TP)	8 (FP)	27
No	7 (FN)	102 (TN)	109
Total	26	110	136

Abbreviations: TP, true positive; FN, false negative; FP, false positive; TN, true negative

Table 3 Performance metrics and confidence intervals.

Metric	Value	95% Confidence Intervals
Accuracy	= 89%	83%–94%
Sensitivity (Recall)	= 0.73	0.52–0.88
Specificity	= 0.93	0.86–0.97
Positive predictive value (PPV)	= 0.70	0.54–0.83
Negative Predictive Value (NPV)	= 0.94	0.89–0.97
Likelihood Ratios for positive test	= 10.05	4.96–20.37
Likelihood Ratios for negative test	= 0.29	0.15–0.55
Blood transfusion prevalence	= 19%	-

Table 4 Screening tool statistics.

Actual Blood Transfusion Group	Precision	Recall	F1-score
Not Received (n = 110)	0.94	0.93	0.93
Received (n = 26)	0.70	0.73	0.72

DISCUSSION

Overall, the accuracy of the prediction model was good (89%) and not significantly different from that of the original model (80%). The model performed well in classifying blood transfusion, as shown by an AUC of 0.83. In the group of patients who did not receive blood transfusion, the prediction model revealed lower precision (0.70) and recall (0.73), indicating a higher rate of misclassification for both FP and FN compared with the group who did not receive blood transfusion. Similarly, the F1-score reasonably indicated a balance between precision and recall; the score for the group that did not receive blood transfusion (0.93) was higher than that of the group that did (0.72).

Regarding the contribution of the model performance, the author's institution is a provincial hospital located outside the capital city. The hospital has limited blood supply in its blood bank. Staffs consistently perform their best work using advanced technology and medical instruments. Life-threatening risks may occur during surgery depending on the patient's condition. Blood transfusion preparation procedures, such as crossmatching 2 units of packed red cells (PRC), remain a crucial routine task that experts agree to maintain, even though the AI prediction model showed high accuracy and precision.

The practical implication of the model performance is as follows: if the model predicts a "YES" result, blood crossmatching of 2 units should be maintained to ensure safety. If the model predicts a "NO" result, the blood crossmatch request for 2 units should be reduced to 1 unit, optimizing cost-effectiveness and benefits, as the surgery is a crucial, life-threatening, elective procedure. As previously stated, real-time monitoring of blood loss, hemoglobin, and hematocrit levels during surgery will determine transfusion needs. According to contingency Table 1, out of 136 cases in this study, only 27 were predicted as "YES" and 109 as "NO." Therefore, the PRC preparation for 109 units could be omitted, optimizing the hospital's blood bank resources[17]. Only 7 cases fell into the FN category; however, 1 unit of PRC was still reserved for each patient. This implication will help balance ambiguity or contradiction. In addition, because of the purposive sampling in this study, the research findings have an acknowledged limitation in generalizability. Regarding the risk factors included in the model, several studies found the same factors in the prediction model, such as gender[13], chronic kidney disease[7,18], ischemic heart disease[7,18], and ASA physical classification[7,18].

Concerning the F1-score, which indicates the balance between precision and recall, these findings reflect a heavily imbalanced dataset (only 20% were classified as the blood transfusion group). As shown in the findings, the precision, recall, and F1-score of the group who did not receive blood transfusions were higher. Regarding research utilization, several factors must be considered.

Recommendations and Limitations

1. Regarding the retrospective and single-center design of this study, there were some significant limitations to data gathering. Incomplete data led to potential bias, resulting in poor predictive model performance. In this study, preoperative hemoglobin concentration, an important high-risk factor that directly influences blood transfusion, was absent from the predictive model. Future research should consider gathering data from prospective designs and multicenter studies to obtain a larger sample size and reduce random error, thereby enhancing model performance. In addition, including more diverse demographic data, strictly maintaining the inclusion and exclusion criteria, or using a higher-power test should be considered to minimize the risk of bias from the research design[19].

2. Some directly significant risk factors, such as preoperative hemoglobin concentration, should be considered for distribution, even though they were not shown to be statistically significant in the variable selection process to be included in the model because of assumption violations.

3. According to the results (Table 4), the predictive model should be considered more suitable for identifying patients who do not require blood transfusion (precision = 0.94) compared with

predicting those who do require blood transfusion (precision = 0.70).

4. Factors such as surgeon volume (high/low) should be considered because a high volume is directly associated with better outcomes. This will improve precision and accuracy[20].

Implication for Clinical Practice

Regarding the implications for clinical practice, the blood transfusion prediction model should be integrated into preoperative workflows to determine whether blood transfusion is required. In cases where the prediction model identifies patients who do not require blood transfusion, the blood transfusion preparation procedure, such as crossmatching 2 units as usual, should be changed to 1 unit to ensure safety during surgery. In cases where the prediction model shows that a transfusion is needed, the blood transfusion preparation procedure should maintain a crossmatch of 2 units as usual.

CONCLUSIONS

Regarding the internal validation in this study, the ML model for blood transfusion probability prediction created by the author was shown to be practically effective in determining the amount of blood required for routine blood transfusion preparation in bipolar hemiarthroplasty. External multicenter validation is recommended for further research to secure advanced conclusions.

REFERENCES

1. Li X, Luo J. Hemiarthroplasty compared to total hip arthroplasty for the treatment of femoral neck fractures: a systematic review and meta-analysis. J Orthop Surg Res 2021;16:172.
2. Bisbe E, Basora M, Colomina MJ. Peri-operative treatment of anaemia in major orthopaedic surgery: a practical approach from Spain. Blood Transfus 2017 ;15:296-306.
3. Coccolini F, Shander A, Ceresoli M, et al. Strategies to prevent blood loss and reduce transfusion in emergency general surgery, WSES-AAST consensus paper. World J Emerg Surg 2024;19:26.
4. Harris JC, Crookston KP. Blood product safety. [Internet]. 2023 [cited 2025 Dec 5]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK539826/.
5. Stoker AD, Binder WJ, Frasco PE, et al. Estimating surgical blood loss: A review of current strategies in various clinical settings. SAGE Open Med 2024;12:20503121241308302.
6. Müller MM, Geisen C, Zacharowski K, et al. Transfusion of packed red cells: indications, triggers and adverse events. Dtsch Arztebl Int 2015;112:507-17; quiz 518.
7. Zang H, Hu A, Xu X, et al. Development of machine learning models to predict perioperative blood transfusion in hip surgery. BMC Med Inform Decis Mak 2024;24:158.
8. Hendrix JM, Garmon EH. American Society of Anesthesiologists Physical Status Classification System. [Internet]. 2025. [cited 2025 Dec 5]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK441940/.
9. Guan J, Li Z, Sheng S, et al. An artificial intelligence-driven revolution in orthopedic surgery and sports medicine. Int J Surg 2025;111:2162-81.
10. İndelen C, Uygun Kızmaz Y, Kar A, et al. The cost of one unit blood transfusion components and cost-effectiveness analysis results of transfusion improvement program. Turk Gogus Kalp Damar Cerrahisi Derg 2021;29:150-7.
11. Liang D, Pang Y, Huang J, et al. Predicting postoperative blood transfusion in elderly patients undergoing total hip and knee arthroplasty using machine learning models. Risk Manag Healthc Policy 2025;18:1697-711.
12. Buddhiraju A, Shimizu MR, Subih MA, et al. Validation of machine learning model performance in predicting blood transfusion after primary and revision total hip arthroplasty. J Arthroplasty 2023;38:1959-66.

13. Luangwaranyoo A, Suksintharanon M, Tangadulrat P, et al. Factors for blood transfusions following hemi hip arthroplasty for patients with femoral neck fracture. Geriatr Orthop Surg Rehabil 2020;22;11: 2151459320972993.
14. Santanapipatkul P. Machine learning risk factors for blood transfusion after hip arthroplasty in fracture neck femur patients. Region 4-5 Medical Journal 2021; 40:381-8.
15. Lalehzarian SP, Gowd AK, Liu JN. Machine learning in orthopaedic surgery. World J Orthop 2021;12:685-99.
16. Hamdi OA, Danan D, Denner A, et al. Impact of implementing stricter criteria for blood transfusion in patients with head and neck cancer undergoing free tissue transfer. Laryngoscope Investig Otolaryngol 2024;9: e70021.
17. Argov S, Shechter Y. Is routine crossmatching for two units of blood necessary in elective surgery? Am J Surg 1981;142:370-1.
18. Seong H, Lee KS, Choi Y, et al. Explainable artificial intelligence for predicting red blood cell transfusion in geriatric patients undergoing hip arthroplasty: Machine learning analysis using national health insurance data. Medicine (Baltimore) 2024;103:e36909.
19. Dziadkowiec O, Durbin J, Muralidharan VJ, et sl. Improving the quality and design of retrospective clinical outcome studies that utilize electronic health records. HCA Healthc J Med 2020;1:131-8.
20. Soldati S, Colais P, Davoli M, et al. More is more? The role of surgeon in the volume-outcome relationship: an Italian population-based cohort study. BMJ Open 2025;15:e098569.